

METASCIENCE
2021 CONFERENCE

Evidence-based steps toward a culture for replicability and reproducibility

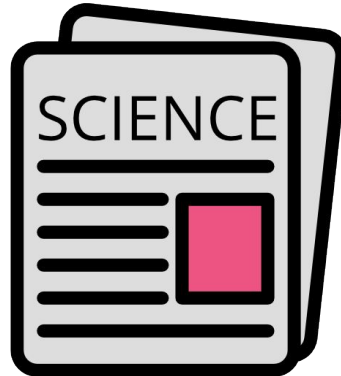
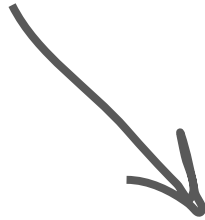
Ana Trisovic

Harvard University





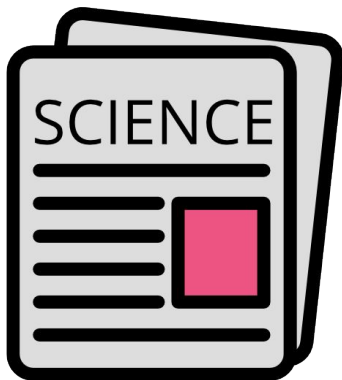
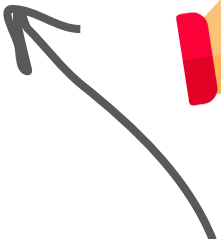
Researchers



Journal



Researchers



Journal

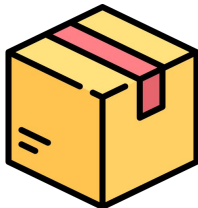


Data repository



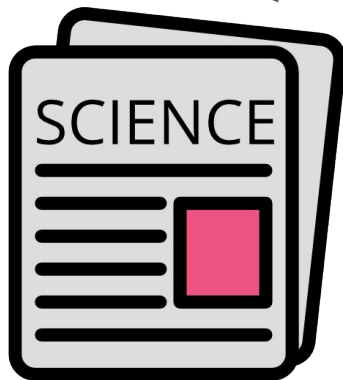
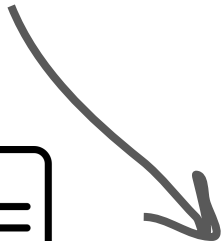
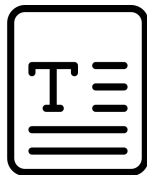
Researchers

Research data
and code

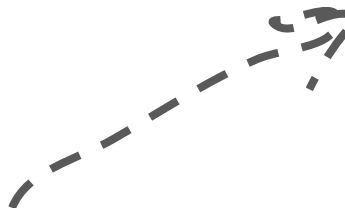


Data repository

Research
paper



Journal



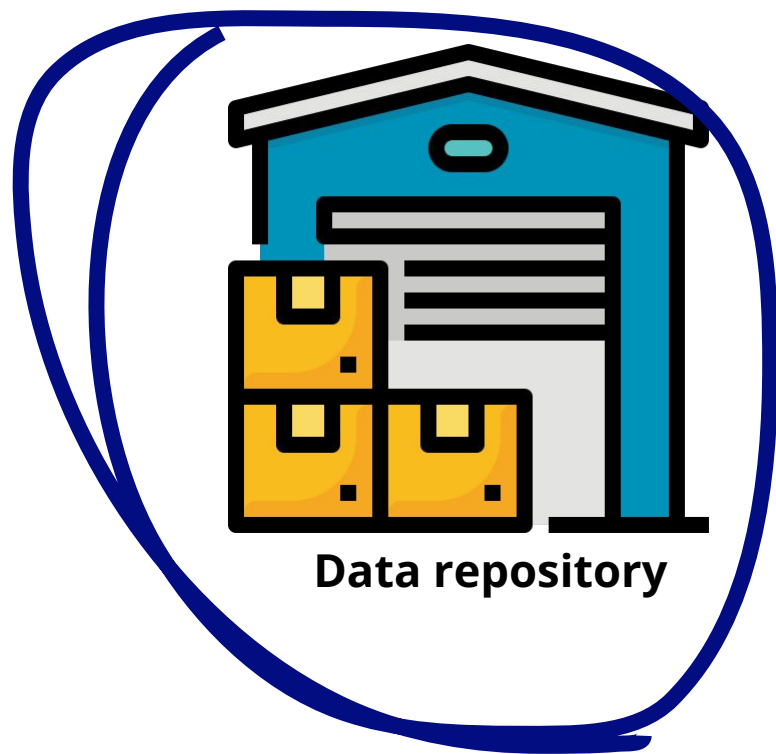
Presentation agenda

1. Dataverse research data repository
2. A large-scale study on code quality and execution
3. Results and discussion
4. What can researchers do
5. What can repositories do
6. What can journals do

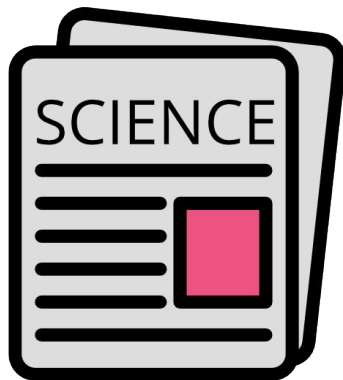
Who are we?



Researchers



Data repository



Journal



- A free and open-source software platform to archive, share, and cite research data
 - Focus on data sharing and making data available
- Provides data repository software that can be installed at institutions
 - Supports research communities for entire countries (NO, NL)

70 institutions around the globe run Dataverse installations as their official data repository



Harvard Dataverse

https://dataverse.harvard.edu

The screenshot shows the Harvard Dataverse website interface. At the top, there is a browser address bar with the URL 'dataverse.harvard.edu'. Below the browser, the website header features the Harvard Dataverse logo on the left and navigation links: 'Add Data', 'Search', 'About', 'User Guide', 'Support', 'Sign Up', and 'Log In'. The main content area is divided into three columns. The first column is titled 'Deposit and share your data. Get academic credit.' and includes a sub-header 'Harvard Dataverse is a repository for research data. Deposit data and code here.' with a button 'Add a dataset +'. The second column is titled 'Organize datasets and gather metrics in your own repository.' and includes a sub-header 'A dataverse is a container for all your datasets, files, and metadata.' with a button 'Add a dataverse +'. The third column is titled 'Publishing your data is easy on Harvard Dataverse!' and includes a sub-header 'Learn about getting started creating your own dataverse repository here.' with a button 'Getting started'. Below these columns, there is a search bar with the text 'Search over 120,600 datasets...' and a 'VIEW ALL DATA >' link. At the bottom, there is a featured section for 'COVID-19 Data Collection', described as 'A curated collection of COVID-19 data deposited in the Harvard Dataverse repository.'

Harvard Dataverse

Add Data Search About User Guide Support Sign Up Log In


Deposit and share your data. Get academic credit.
Harvard Dataverse is a repository for research data. Deposit data and code here.
[Add a dataset +](#)

Organize datasets and gather metrics in your own repository.
A dataverse is a container for all your datasets, files, and metadata.
[Add a dataverse +](#)

Publishing your data is easy on Harvard Dataverse!
Learn about getting started creating your own dataverse repository here.
[Getting started](#)

Find data across research fields, preview metadata, and download files

Search over 120,600 datasets... [VIEW ALL DATA >](#)

Featured  **COVID-19 Data Collection**
A curated collection of COVID-19 data deposited in the Harvard Dataverse repository.

Data sharing

- Stand-alone or institutional account for depositing data (Dataverse, GitHub, Google, ORCID, University credentials)
- Individuals, institutes or journals may have own dataverse collections

The screenshot shows the Harvard Dataverse submission interface. At the top, the Harvard Dataverse logo is on the left, and navigation links for 'Add Data', 'Search', 'About', 'User Guide', and 'Support' are on the right. The main section is titled 'Host Dataverse' and contains a text input field with 'Harvard Dataverse' entered. Below this is a note: 'Changing the host dataverse will clear any fields you may have entered data into.' A warning message states '*Asterisks indicate required fields'. The 'Citation Metadata' section is expanded and contains several fields: 'Title' (required), 'Author' (required), 'Contact' (required), 'Description' (required), 'Subject' (required), 'Keyword' (required), and 'Related Publication' (required). Each field has a help icon. The 'Author' and 'Contact' sections are populated with 'Trisovic, Ana' and 'anatrivic@fas.harvard.edu' respectively. The 'Description' field has a note about HTML tags. The 'Subject' field is a dropdown menu. The 'Keyword' section includes 'Term', 'Vocabulary', and 'Vocabulary URL' fields. The 'Related Publication' field is currently empty.

HARVARD
Dataverse

Add Data - Search - About User Guide Support

Host Dataverse ⓘ Changing the host dataverse will clear any fields you may have entered data into.
Harvard Dataverse

*Asterisks indicate required fields

Citation Metadata ^

Title * ⓘ Enter title...
Add "Replication Data for" to Title

Author * ⓘ **Name** * ⓘ Trisovic, Ana **Affiliation** ⓘ Harvard University +
Identifier Scheme ⓘ Select... **Identifier** ⓘ

Contact * ⓘ **Name** ⓘ Trisovic, Ana **Affiliation** ⓘ Harvard University +
E-mail * ⓘ anatrivic@fas.harvard.edu

Description * ⓘ This field supports only certain [HTML tags](#). **Text** * ⓘ +

Subject * ⓘ Select...
Keyword ⓘ **Term** ⓘ **Vocabulary** ⓘ +
Vocabulary URL ⓘ Enter full URL, starting with http://

Related Publication ⓘ **Citation** ⓘ +

Data sharing

- Stand-alone or institutional account for depositing data (Dataverse, GitHub, Google, ORCID, University credentials)
- Individuals, institutes or journals may have own dataverse collections

The screenshot shows the Harvard Dataverse data entry interface. At the top, the Harvard Dataverse logo is on the left, and navigation links for 'Add Data', 'Search', 'About', 'User Guide', and 'Support' are on the right. The main section is titled 'Host Dataverse' and contains a text input field with 'Harvard Dataverse' entered. Below this is a note: 'Changing the host dataverse will clear any fields you may have entered data into.' A warning message states '*Asterisks indicate required fields'. The 'Citation Metadata' section is expanded and contains several fields: 'Title' (required), 'Author' (required), 'Contact' (required), 'Description' (required), 'Subject' (required), 'Keyword' (required), and 'Related Publication' (required). Each field has a help icon. The 'Author' and 'Contact' fields are grouped and include sub-fields for 'Name', 'Affiliation', and 'E-mail'. The 'Description' field has a note about HTML tags. The 'Keyword' field includes 'Term' and 'Vocabulary' sub-fields. The 'Related Publication' field includes a 'Citation' sub-field. Plus signs are visible to the right of several fields, indicating they can be added or expanded.

HARVARD
Dataverse

Add Data - Search - About User Guide Support

Host Dataverse ⓘ Changing the host dataverse will clear any fields you may have entered data into.

Harvard Dataverse

*Asterisks indicate required fields

Citation Metadata ^

Title * ⓘ Enter title...
Add "Replication Data for" to Title

Author * ⓘ

Name * ⓘ Trisovic, Ana
Affiliation ⓘ Harvard University +

Identifier Scheme ⓘ Select...
Identifier ⓘ

Contact * ⓘ

Name ⓘ Trisovic, Ana
Affiliation ⓘ Harvard University +

E-mail * ⓘ anatrivic@fas.harvard.edu

Description * ⓘ This field supports only certain HTML tags.
Text * ⓘ +

Date ⓘ YYYY-MM-DD

Subject * ⓘ Select...
Keyword * ⓘ

Term ⓘ **Vocabulary** ⓘ +

Vocabulary URL ⓘ Enter full URL, starting with http://

Related Publication ⓘ

Citation ⓘ +

Replication dataset

- Replication dataset - a bundle of data, code and other files needed to reproduce a published study

HARVARD Dataverse Add Data Search About

POLITICAL ANALYSIS

Political Analysis Dataverse (Cambridge University Press)

HARVARD Dataverse Add Data Search About User Guide Support

AJPS AMERICAN JOURNAL of POLITICAL SCIENCE

American Journal of Political Science (AJPS) Dataverse (Midwest Political Science Association) ajps.org

Harvard Dataverse > American Journal of Political Science (AJPS) Dataverse

The *American Journal of Political Science* is committed to significant advances in knowledge and understanding of citizenship, gov politics, and to the public value of political science research. To find out more about our data integrity policies, please visit our web

Replication Data for: How Political Parties Shape Public Opinion in the Real World

Version 2.0



Bisgaard, Martin; Rune Slothuus, 2020, "Replication Data for: How Political Parties Shape Public Opinion in the Real World", <https://doi.org/10.7910/DVN/Z5BTCQ>, Harvard Dataverse, V2, UNF-6:YTyX+kjbxSUNEND/3GGg== [fileUNF]

[Cite Dataset](#)

[Learn about Data Citation Standards.](#)

Access Dataset -
Contact Owner Share

Dataset Metrics
1,092 Downloads

Description

How powerful are political parties in shaping citizens' opinions? Despite longstanding interest in the flow of influence between partisan elites and citizens, few studies to date examine how citizens react when their party changes its position on a major issue in the real world. We present a rare quasi-experimental panel study of how citizens responded when their political party suddenly reversed its position on two major and salient welfare issues in Denmark. With a five-wave panel survey collected just around these two events, we show that citizens' policy opinions changed immediately and substantially when their party switched its policy position—even when the new position went against citizens' previously held views. These findings advance the current, largely experimental literature on partisan elite influence. (2020-03-26)

Subject

Social sciences

Keyword

Party cues, Political parties, Elite influence, Motivated reasoning, Polarization, Public opinion, Panel survey

Related Publication

Bisgaard, Martin, and Rune Slothuus. [date]. "How Political Parties Shape Public Opinion in the Real World." *American Journal of Political Science* Forthcoming. <http://ajps.org/>

Notes

This dataset underwent an independent verification process that replicated the tables and figures in the primary article. For the supplementary materials, verification was performed solely for the successful execution of code. The verification process was carried out by the Odum Institute for Research in Social Science at the University of North Carolina at Chapel Hill.

The associated article has been awarded Open Materials and Open Data Badges. Learn more about the Open Practice Badges from the Center for Open Science.



Files Metadata Terms Versions

Search this dataset...

Filter by
File Type: All Access: All

Sort

1 to 10 of 25 Files Download

	build_data.R R Syntax - 12.1 KB Published Jun 29, 2020 56 Downloads MD5: a94...597		
	codebook_ess.pdf Adobe PDF - 508.8 KB Published Jun 29, 2020 46 Downloads		

Replication dataset

- Replication dataset - a bundle of data, code and other files needed to reproduce a published study

Replication Data for: How Political Parties Shape Public Opinion in the Real World

Version 2.0

Bisgaard, Martin; Rune Slothuus, 2020, "Replication Data for: How Political Parties Shape Public Opinion in the Real World", <https://doi.org/10.7910/DVN/Z5BTCQ>, Harvard Dataverse, V2, UNF-6:YTyX+kjbsSZUNEND/3GGg== [fileUNF]

[Cite Dataset](#)

[Learn about Data Citation Standards.](#)

[Access Dataset](#)

[Contact Owner](#) [Share](#)

[Dataset Metrics](#)

1,092 Downloads

Dataset metrics

Description

How powerful are political parties in shaping citizens' opinions? Despite longstanding interest in the flow of influence between partisan elites and citizens, few studies to date examine how citizens react when their party changes its position on a major issue in the real world. We present a rare quasi-experimental panel study of how citizens responded when their political party suddenly reversed its position on two major and salient welfare issues in Denmark. With a five-wave panel survey collected just around these two events, we show that citizens' policy opinion immediately and substantially when their party switched its policy position—even if their position went against citizens' previously held views. These findings advance the experimental literature on partisan elite influence. (2020-03-26)

Subject

Social Sciences

Keyword

Party cues, Political parties, Elite influence, Motivated reasoning, Polarization, Public opinion, Panel survey

Related Publication

Bisgaard, Martin, and Rune Slothuus. [date]. "How Political Parties Shape Public Opinion in the Real World." *American Journal of Political Science* Forthcoming. <http://ajps.org/>

Notes

This dataset underwent an independent verification process that replicated the tables and figures in the primary article. For the supplementary materials, verification was performed solely for the successful execution of code. The verification process was carried out by the Odum Institute for Research in Social Science at the University of North Carolina at Chapel Hill.

The associated article has been awarded Open Materials and Open Data Badges. Learn more about the Open Practice Badges from the [Center for Open Science](#).



Center for Open Science Badges

[Files](#) [Metadata](#) [Terms](#) [Versions](#)

Search this dataset...

Filter by
File Type: All - Access: All -

1 to 10 of 25 Files

build_data.R
R Syntax - 12.1 KB
Published Jun 29, 2020
56 Downloads
MD5: a04...597

Code, documentation and other files

codebook_ess.pdf
Adobe PDF - 508.8 KB
Published Jun 29, 2020
46 Downloads



Summary

- Dataverse data repositories have versatile support for data sharing
- Research data and code are shared in a “replication dataset” that often belong to a journal or institutional collection

How reusable are our replication datasets?



A large-scale study on research code quality and re-execution

**A large-scale
study on research**

**code quality and
re-execution**



**reusable
replication
datasets**

A large-scale study on research code quality and re-execution



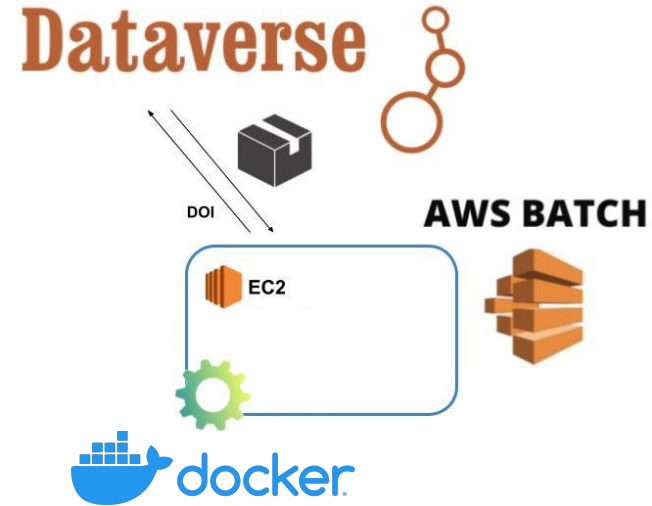
reusable
replication
datasets



not reusable
replication
datasets

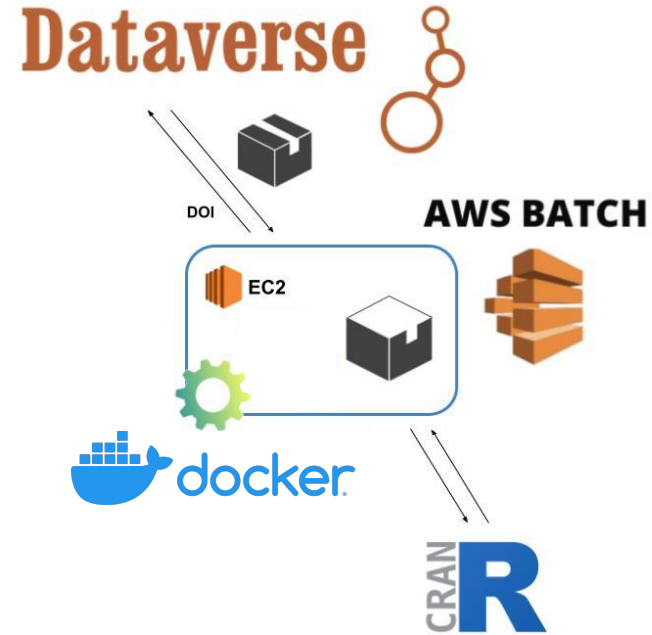
Data collection workflow

1. Replication dataset is retrieved from Harvard Dataverse to AWS
2. We collect data on the content, code, install used libraries etc.
3. We attempt code re-execution for an allocated time of 1h per file and 5h in total
4. The re-execution result and other collected data are passed to the backend database for analysis



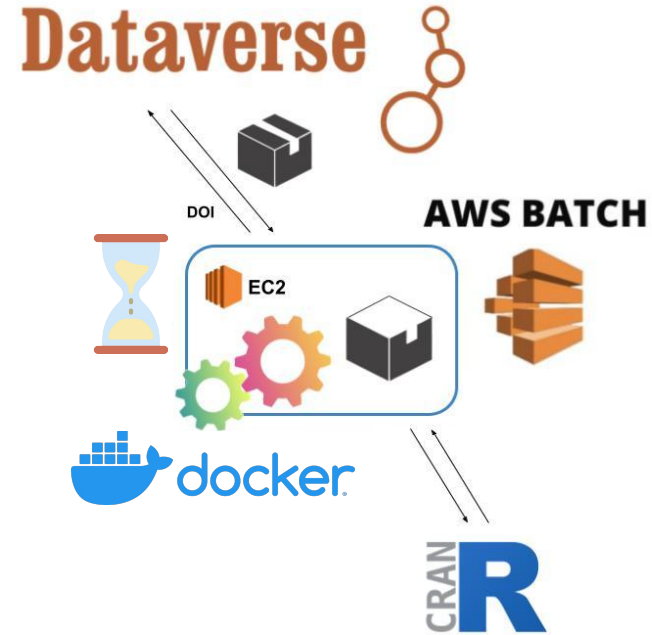
Data collection workflow

1. Replication dataset is retrieved from Harvard Dataverse to AWS
2. We collect data on the content, code, install used libraries etc.
3. We attempt code re-execution for an allocated time of 1h per file and 5h in total
4. The re-execution result and other collected data are passed to the backend database for analysis



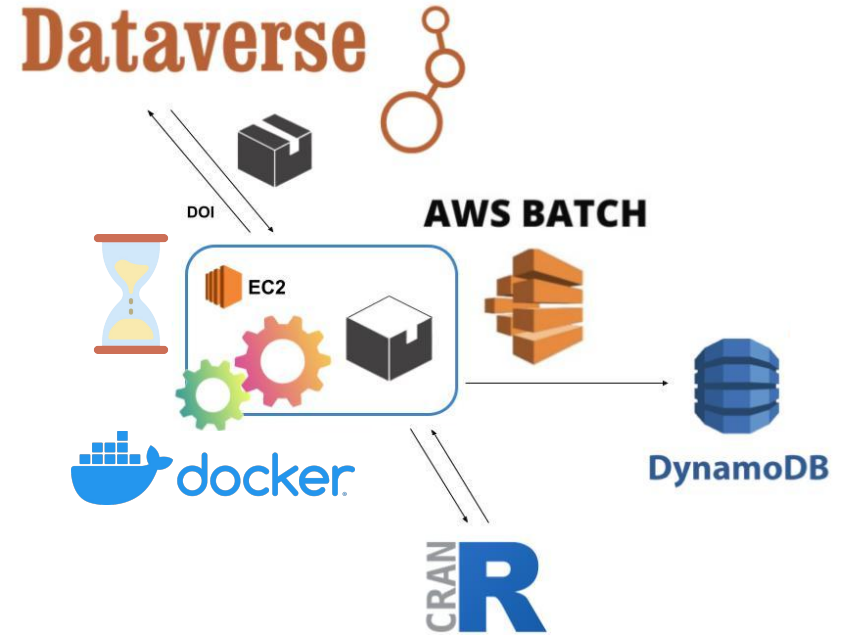
Data collection workflow

1. Replication dataset is retrieved from Harvard Dataverse to AWS
2. We collect data on the content, code, install used libraries etc.
3. We attempt code re-execution for an allocated time of 1h per file and 5h in total
4. The re-execution result and other collected data are passed to the backend database for analysis



Data collection workflow

1. Replication dataset is retrieved from Harvard Dataverse to AWS
2. We collect data on the content, code, install used libraries etc.
3. We attempt code re-execution for an allocated time of 1h per file and 5h in total
4. The re-execution result and other collected data are passed to the backend database for analysis



Results (basic properties)

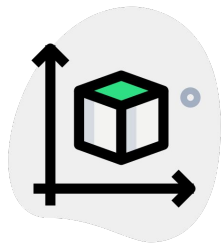
Retrieved 2109 publicly
available replication datasets
containing 9078 R files

Over 94% of the datasets
belonged to social sciences

Results (basic properties)

Retrieved 2109 publicly
available replication datasets
containing 9078 R files

Over 94% of the datasets
belonged to social sciences



Dataset size (median):
3.2 MB

Number of files (median):
8 (typically less than 15)

Results (basic properties)

Retrieved 2109 publicly available replication datasets containing 9078 R files

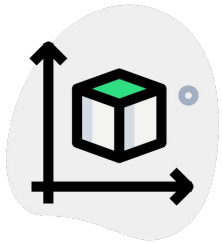
Over 94% of the datasets belonged to social sciences

Dataset size (median):
3.2 MB

Number of files (median):
8 (typically less than 15)



- File name length: 10-20 characters
- Documentation present in 57% of the datasets
- Comments comprise 20% of the shared R code



Presence of conventional files

DESCRIPTION

README.md

LICENSE

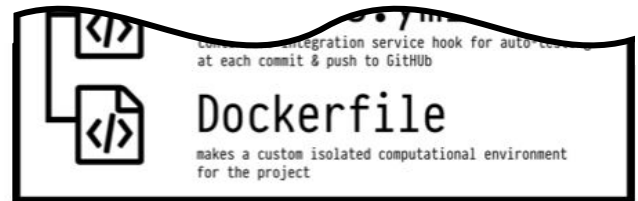
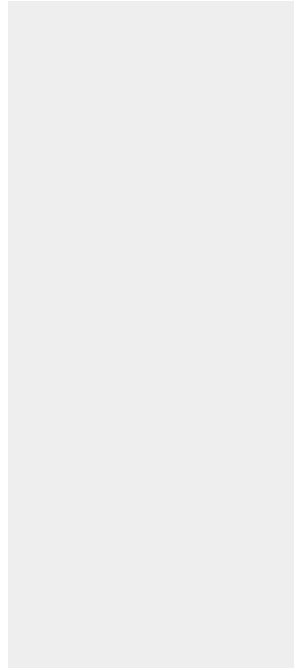
NAMESPACE

Dockerfile

R Markdown

.Rproj

install.R



Presence of conventional files

DESCRIPTION

< 1%

README.md

LICENSE

< 1%

NAMESPACE

< 1%

Dockerfile

< 1%

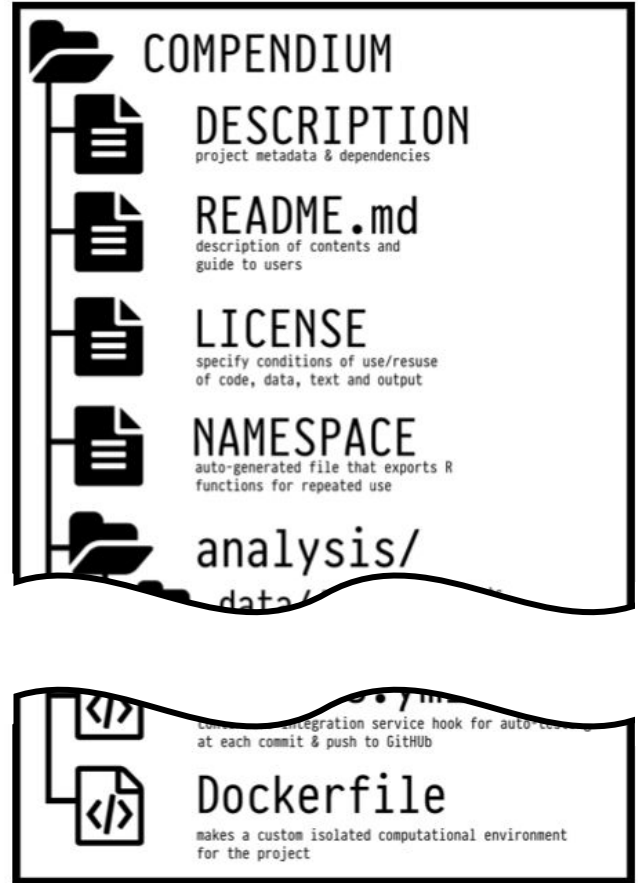
R Markdown

.Rproj

< 1%

install.R

< 1%



Presence of conventional files

DESCRIPTION	< 1%
README.md	48%
LICENSE	< 1%
NAMESPACE	< 1%
Dockerfile	< 1%
R Markdown	3%
.Rproj	< 1%
install.R	< 1%



Out of 2109
replication
datasets

Most used libraries in research code:



1. **Data visualization and plotting** (`ggplot2`,
`lattice`)



2. **Data wrangling and display in a tabular form** (`xtable`)



3. **Data import and export** (`foreign`, `dplyr`,
`plyr`, `reshape2`)



4. **Statistical analysis** (`stargazer`, `MASS`,
`lmtest`, `car`)

In the research code, no libraries detected for:



1. **Code testing** (`runit`, `testthat`,
`tinytest`, `unitizer`)



2. **Provenance tracking** (`provR`,
`provenance`, `RDTlite`, `provTraceR`)

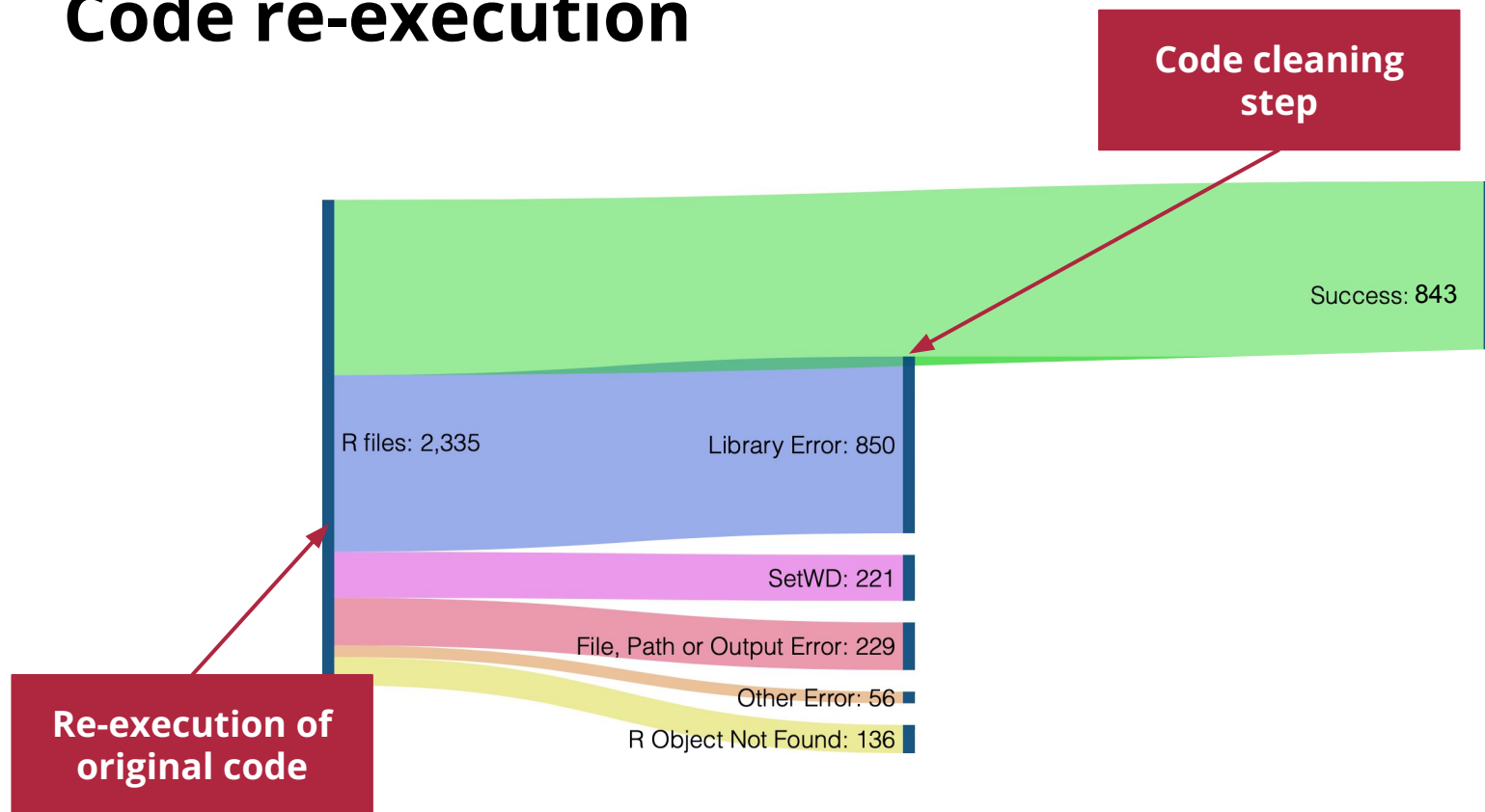


3. **Runtime environment management**
(`packrat`, `pacman`)

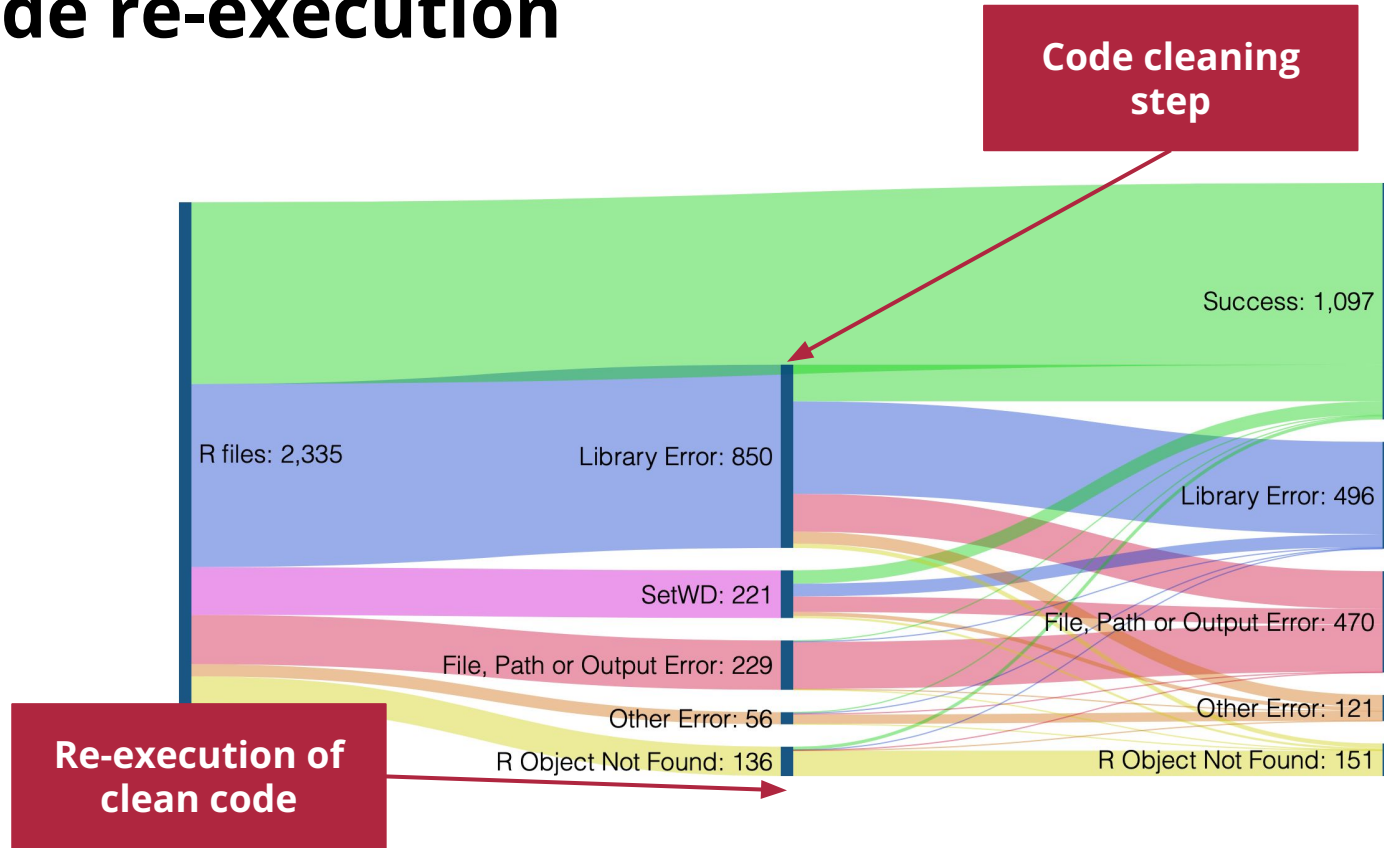


4. **Workflow libraries** (`workflowR`,
`workflows`, `drake`)

Code re-execution

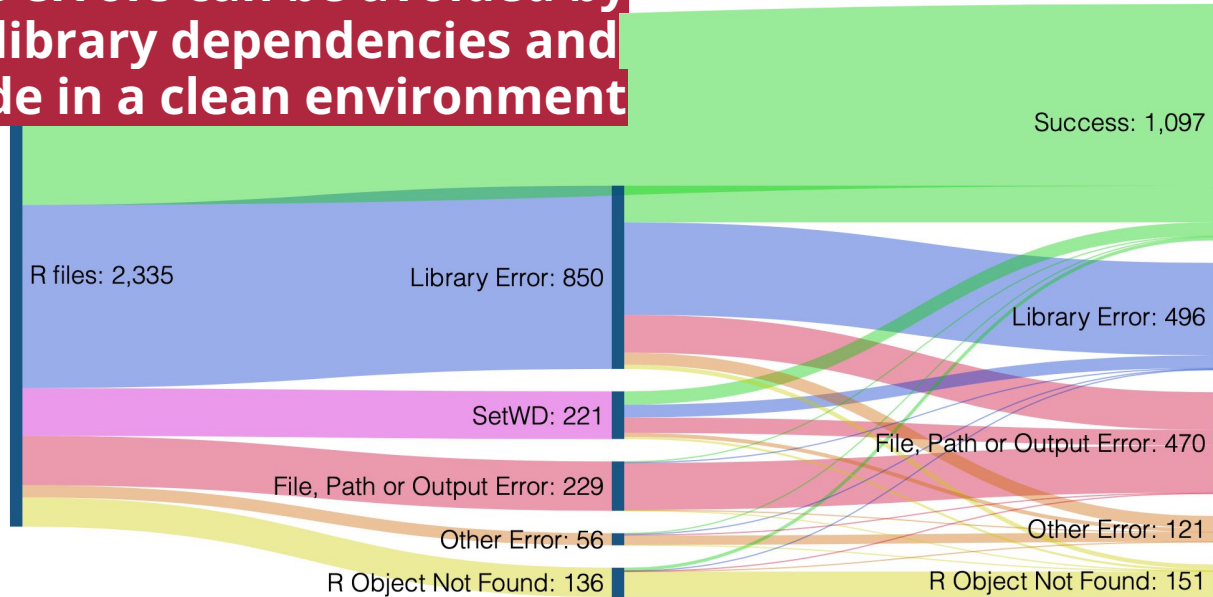


Code re-execution



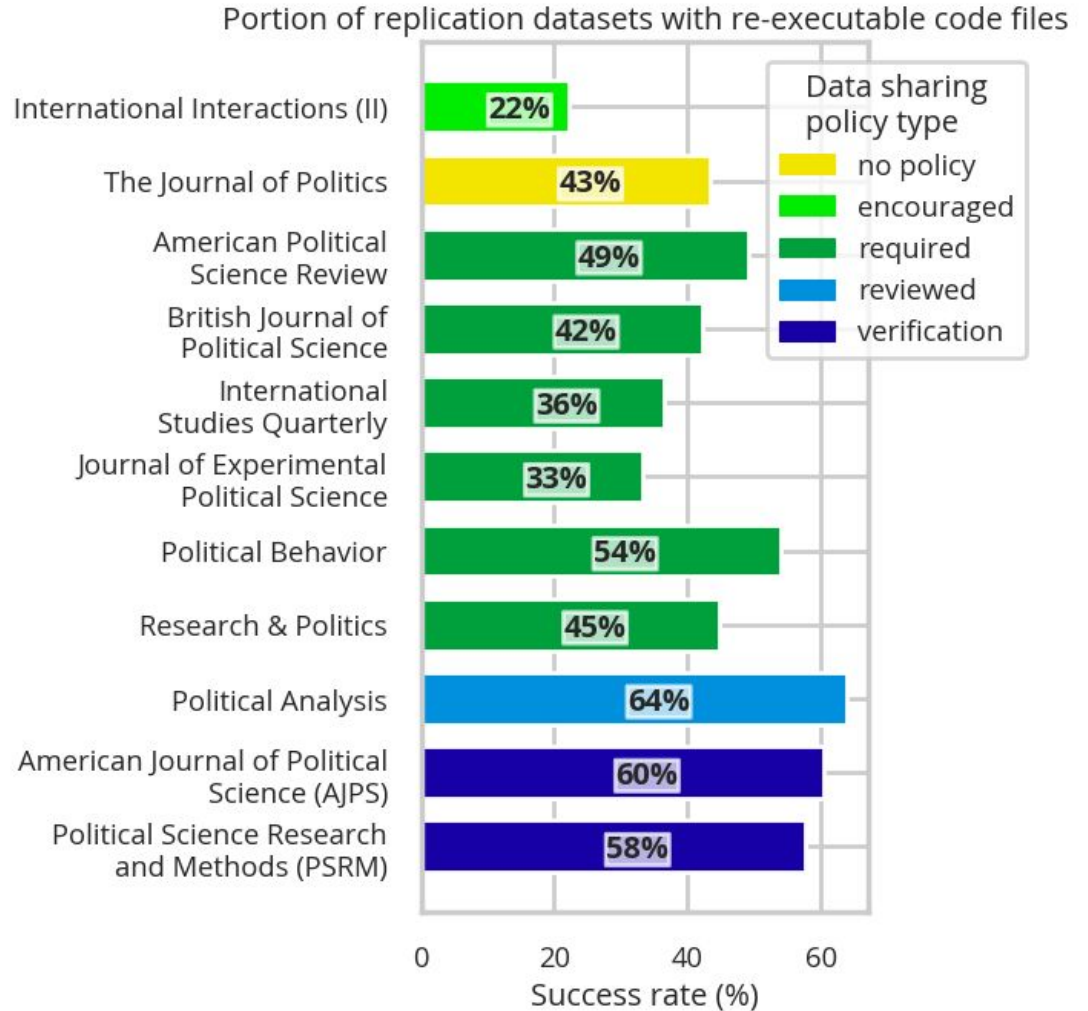
Code re-execution

Many code errors can be avoided by capturing library dependencies and testing code in a clean environment



Journals with stricter data policies have higher rate of executable code

Journal average: 47%
Total average: 45%



Summary

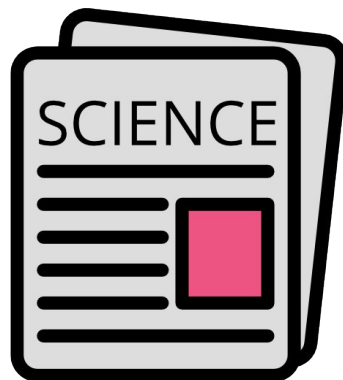
- No libraries for code testing, provenance or workflows
- Simple code cleaning resulted in substantial improvement in re-execution
- Re-execution correlates with journals' data sharing policy strictness



Researchers



Data repository



Journal

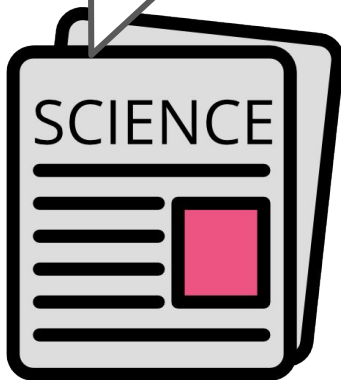


Researchers

We want open
and reproducible
science



Data repository



Journal

**What can
researchers do?**



Best practices and recommendations

1. Library versions should be captured by, minimally, the output of `sessionInfo()` from the researcher's R session, or a DESCRIPTION file, or `install.R`, or by using the `renv` package to track the libraries and their version number.
2. When referring to data, code or other files, use relative file paths, as full paths will cause an error when the code is executed on other systems.

Best practices and recommendations

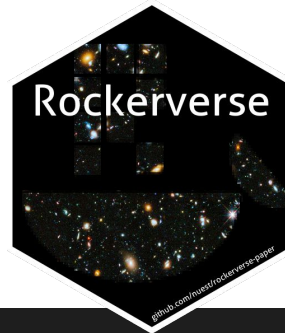
1. Library versions should be captured by, minimally, the output of `sessionInfo()` from the researcher's R session, or a DESCRIPTION file, or `install.R`, or by using the `renv` package to track the libraries and their version number.
2. When referring to data, code or other files, use relative file paths, as full paths will cause an error when the code is executed on other systems.

Best practices and recommendations

3. Workflow capture and management methods such as R Markdown, targets (or drake) will help to automate your code and specify the correct execution sequence.
4. Use Docker to document your runtime environment in a machine-readable format, and to ensure others can recreate your computing environment.

Best practices and recommendations

3. Workflow capture and management methods such as R Markdown, targets (or drake) will help to automate your code and specify the correct execution sequence.
4. Use Docker to document your runtime environment in a machine-readable format, and to ensure others can recreate your computing environment.

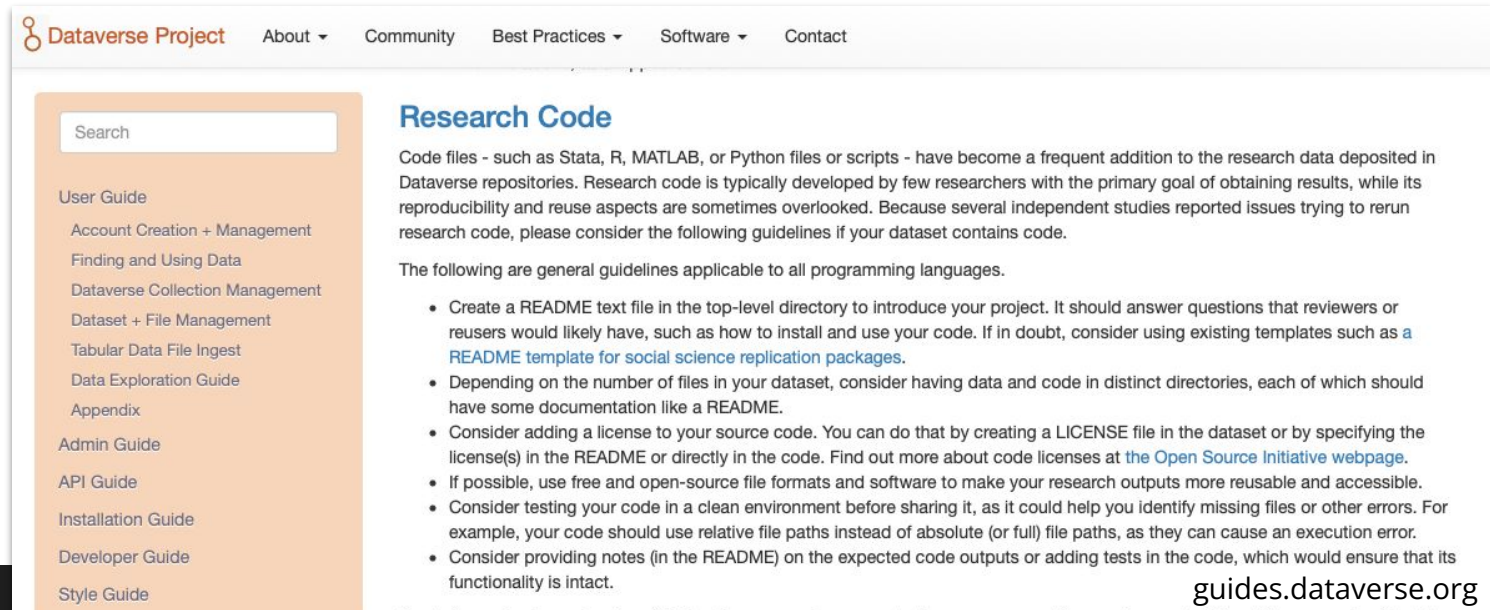


**What can
repositories do?**



Best practices and recommendations

1. Create and maintain documentation on adequate deposit of research code.



The screenshot shows the Dataverse Project website. The navigation bar includes 'Dataverse Project', 'About', 'Community', 'Best Practices', 'Software', and 'Contact'. A search bar is located on the left. The main content area is titled 'Research Code' and contains the following text:

Code files - such as Stata, R, MATLAB, or Python files or scripts - have become a frequent addition to the research data deposited in Dataverse repositories. Research code is typically developed by few researchers with the primary goal of obtaining results, while its reproducibility and reuse aspects are sometimes overlooked. Because several independent studies reported issues trying to rerun research code, please consider the following guidelines if your dataset contains code.

The following are general guidelines applicable to all programming languages.

- Create a README text file in the top-level directory to introduce your project. It should answer questions that reviewers or reusers would likely have, such as how to install and use your code. If in doubt, consider using existing templates such as [a README template for social science replication packages](#).
- Depending on the number of files in your dataset, consider having data and code in distinct directories, each of which should have some documentation like a README.
- Consider adding a license to your source code. You can do that by creating a LICENSE file in the dataset or by specifying the license(s) in the README or directly in the code. Find out more about code licenses at [the Open Source Initiative webpage](#).
- If possible, use free and open-source file formats and software to make your research outputs more reusable and accessible.
- Consider testing your code in a clean environment before sharing it, as it could help you identify missing files or other errors. For example, your code should use relative file paths instead of absolute (or full) file paths, as they can cause an execution error.
- Consider providing notes (in the README) on the expected code outputs or adding tests in the code, which would ensure that its functionality is intact.

guides.dataverse.org

Best practices and recommendations

2. Integrations with reproducibility platforms such as CodeOcean, WholeTale, Jupyter Binder and Renku will facilitate environment capture and encapsulation of research code.



Best practices and recommendations

3. An internal working group will help identify community-wide problems, prioritize them, and implement solutions.



GDCC

The Global Dataverse Community Consortium
Supporting Dataverse repositories Around the World

GDCC Software, Workflows & Containers Working Group

Members: Oliver Bertuch, Danny Brooke, Philip Durbin, Julian Gautier, Dorothea Iglezakis, Genevieve Michaud, Katie Mika, Jim Myers, Slava Tykhonov, Tom Vilette

Chair: [Ana Trisovic](#)

Motivation and goals

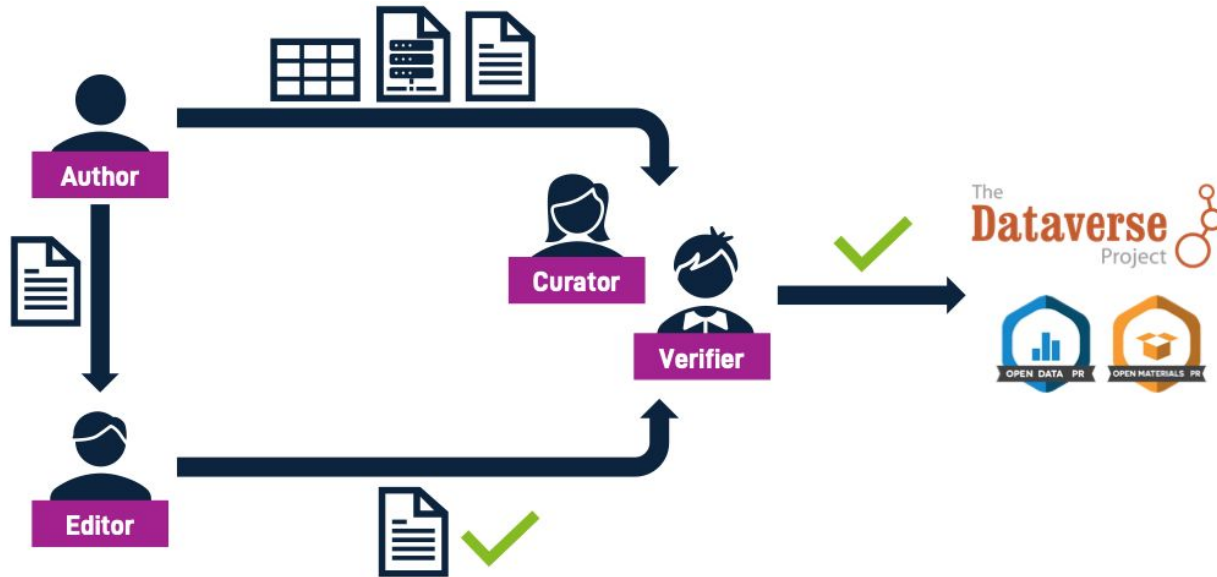
The SWC group aims to support research software, workflows, and container dissemination for reproducibility and reuse. The group discusses the necessary metadata, file formats, tools, and infrastructure necessary to incorporate these resources in data repositories

**What can
journals do?**



Best practices and recommendations

1. Encourage a simple review of all deposited material if a code verification is infeasible



@amagooch

Best practices and recommendations

2. Create reproducibility checklist or templates for authors



A template README for social science replication packages.

The template README provided on this website is in a form that follows best practices as defined by a number of data editors at social science journals.

<https://social-science-data-editors.github.io>

Template README and Guidance

INSTRUCTIONS: This README suggests structure and content that have been approved by various journals, see [Endorsers](#). It is available as [Markdown/txt](#), [Word](#), [LaTeX](#), and [PDF](#). In practice, there are many variations and complications, and authors should feel free to adapt to their needs. All instructions can (should) be removed from the final README (in Markdown, remove lines starting with > `INSTRUCTIONS`). Please ensure that a PDF is submitted in addition to the chosen native format.

Overview

INSTRUCTIONS: The typical README in social science journals serves the purpose of guiding a reader through the available material and a route to replicating the results in the research paper. Start by providing a brief overview of the available material and a brief guide as to how to proceed from beginning to end.

Example: The code in this replication package constructs the analysis file from the three data sources (Ruggles et al, 2018; Inglehart et al, 2019; BEA, 2016) using Stata and Julia. Two main files run all of the code to generate the data for the 15 figures and 3 tables in the paper. The replicator should expect the code to run for about 14 hours.

Data Availability and Provenance Statements

@larsvil

Best practices and recommendations

3. Integrations with reproducibility platforms



**eLife and Stencila announce roadmap
for bringing reproducible publishing to
more authors**

The next phase of the Executable Research Article project will focus on reducing barriers to the authoring and publication of reproducible research papers.

[Home](#) > [arXiv updates](#) > Instant access to code, for any arXiv paper

Instant access to code, for any arXiv paper



Papers With Code

partners with

arXiv.org

Conclusions

Conclusion

- We've seen evidence of both good and bad coding and dissemination practices (documentation, commenting, convention files rarely used)
- It is hard to re-execute "old" code and even harder to reuse it
 - Curated replication datasets have higher re-execution rates.
 - Things are looking up!
- Employing proposed recommendations would help researchers, repositories and journals contribute to research transparency and reproducibility.

Conclusion

- We've seen evidence of both good and bad coding and dissemination practices (documentation, commenting, convention files rarely used)
- It is hard to re-execute "old" code and even harder to reuse it
 - Curated replication datasets have higher re-execution rates.
 - Things are looking up!
- Employing proposed recommendations would help researchers, repositories and journals contribute to research transparency and reproducibility.

References

- This presentation was based on findings at Trisovic, Ana, et al. "A large-scale study on research code quality and execution." arXiv:2103.12793 (2021).

Acknowledgements:
icons by Freepik, Pixel perfect, Nhor Phai, Darius Dan (flaticon.com)

A large-scale study on research code quality and execution

Ana Trisovic^{1*}, Matthew K. Lau², Thomas Pasquier³, Mercè Crosas¹

March 25, 2021

1. Institute for Quantitative Social Science, Harvard University, Cambridge, MA, USA
2. CAS Key Laboratory of Forest Ecology and Management, Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang, China
3. Department of Computer Science, University of British Columbia, Vancouver, BC, Canada

*corresponding author(s): Ana Trisovic (anatrivic@g.harvard.edu)

Abstract

This article presents a study on the quality and execution of research code from publicly-available replication datasets at the Harvard Dataverse repository. Research code is typically created by a group of scientists and published together with academic papers to facilitate research transparency and reproducibility. For this study, we define ten questions to address aspects impacting research reproducibility and reuse. First, we retrieve and analyze more than 2000 replication datasets with over 9000 unique R files published from 2010 to 2020. Second, we execute the code in a clean runtime environment to assess its ease of reuse. Common coding errors were identified, and some of them were solved with automatic code cleaning to aid code execution. We find that 74% of R files crashed in the initial execution, while 56% crashed when code cleaning was applied, showing that many errors can be prevented with good coding practices. We also analyze the replication datasets from journals' collections and discuss the impact of the journal policy strictness on the code re-execution rate. Finally, based on our results, we propose a set of recommendations for code dissemination aimed at researchers, journals, and repositories.

1 Introduction

Researchers increasingly publish their data and code to enable scientific transparency, reproducibility, reuse, or compliance with funding bodies, journals, and academic institutions [1]. Reusing data and code should propel new research and save researchers' time, but in practice, it is often easier to write new code



ALFRED P. SLOAN
FOUNDATION



Thank you!

Email: anatrisovic@g.harvard.edu
GitHub & Twitter: [atrisovic](#)
Dataverse Project: <https://dataverse.org/contact>